

METHOD FOR THE IDENTIFICATION AND SPECIATION OF
BACTERIA OF THE *BURKHOLDERIA* CEPACIA COMPLEX

Field of the Invention

This application relates to a new method for the identification and speciation of bacteria of the *Burkholderia cepacia* complex, and to vaccines developed specific for certain bacteria characterized using this method.

Background of the Invention

The Gram negative bacterium *Burkholderia cepacia* has recently been shown to consist of five different genomovars or new species, and as a collective the bacteria have been called the *B. cepacia* complex. Two of these genomovars have been given new species names: *Burkholderia multivorans* (formerly genomovar II) and *Burkholderia vietnamiensis* (formerly genomovar V). These bacteria cause problematic infections in patients with cystic fibrosis (CF) and chronic granulomatous disease as well as often causing infection outbreaks among vulnerable hospitalized patients. In cystic fibrosis, clinical outcome and epidemiology of infection may vary depending on the type of species patients are colonized with. It has become critical to obtain a rapid and reproducible means of identifying the different species of the *B. cepacia* complex since the current genomovar classification is technically difficult, laborious, carried out by only one reference laboratory and not based on a single test. Vandamme et al., *Int'l. J. Systematic Bacteriol.* 1188-1200 (1997).

It is an object of the present invention to provide a rapid and reproducible method for identification and speciation of *B. cepacia* complex.

It is a further object of the invention to provide suitable reagents and kits for use in the method of the invention.

Using this test, or other established but more time consuming tests, the species of a bacteria of the *B. cepacia* complex which is responsible for a given infection may be determined. There remains, however, the challenge of provide effective therapy if the strain is determined to be an epidemic strain. It is therefore still a further object of the present invention to provide an answer to this challenge by providing a vaccine which

- 2 -

promotes a therapeutically beneficial immune response to epidemic strains of the *B. cepacia*.

Summary of The Invention

In establishing a methodology for identification and speciation of a target group of microorganisms based on nucleotide sequences, it is necessary to identify a gene or genes within the target group which on the one hand contain conserved regions which are common to all microorganisms in the group such that non-specific amplification can be performed, and on the other hand contain regions which are dissimilar in ways that are diagnostic of the species. In accordance with the invention, it has been determined that these characteristics are found in the *recA* gene of bacteria of the *Burkholderia cepacia* complex. Thus, in one aspect of the invention, identification and speciation of bacteria of the *Burkholderia cepacia* complex in a sample can be accomplished by a method comprising the steps of

- (a) obtaining nucleotide sequence information for the *recA* gene in bacteria of the *Burkholderia cepacia* complex found in the sample; and
- (b) comparing the nucleotide sequence information obtained for the *recA* gene in bacteria of the *Burkholderia cepacia* complex found in the sample with a standard library of nucleotide sequence information comprising standard nucleotide sequence information for at least three species of bacteria of the *Burkholderia cepacia* complex.

Preferably, the nucleotide sequence information is obtained by evaluation of restriction fragment length polymorphism (RFLP). Other techniques for obtaining sequence information can also be used, including base-by-base determination of the sequence of the region of interest, sequence-specific oligonucleotide hybridization probes, and ligation techniques.

The invention also provides universal primers which can be used for amplification of all known members of the *Burkholderia cepacia* complex, and genomovar-specific primers which can be used for selective amplification of the *recA* gene from bacteria of one genomovar.

Speciation of bacteria of the *Burkholderia cepacia* complex can have several important consequences. When an infection has been identified as being caused by an

- 3 -

epidemic strain of the *B. cepacia* complex, it is appropriate to provide a vaccine which can be used in treating and preventing infection. Thus, the present invention also provides a vaccine composition which is useful in providing a therapeutically beneficial immune response in warm-blooded animals, particularly mammals (including humans) infected with epidemic strains of bacteria of the *B. cepacia* complex belonging to genomovar III. The vaccine comprises a protein or peptide antigen, or an expressible polynucleotide encoding a protein or peptide antigen derived from the flagellin of such bacteria. When the vaccine composition is administered to a mammalian subject, a therapeutically beneficial immune response is stimulated, which assists in combating the *B. cepacia* infection.

Speciation can also be used as a basis for selection and/or isolation of industrially useful bacterial species. For example, it is known that a number of strains of *B. cepacia* can be used in as biocontrol strains in agricultural applications. It has been found using the speciation method of the invention that these strains frequently fall into one of two classes based on *recA* RFLP pattern, one of which is a class associated with epidemic strains and one of which is not. Thus, one can use the test methodology of the invention to select among biocontrol strains to reduce the likelihood of public health hazards.

Brief Description of The Drawings

Figs. 1 and 2 show phylogenetic trees for *B. cepacia* based on the *recA* sequence evaluation; and

Fig. 3 shows the RFLP patterns obtained for representative strains of *B. cepacia* complex.

Detailed Description of The Invention

DNA sequence variation in conserved bacterial genes, such as those encoding the ribosomal RNA (rRNA) molecules, has been widely used as a method for classification and speciation of bacteria. Speciation of *B. cepacia* complex based on the genes encoding the 16S rRNA gene has been examined, but there is insufficient DNA variation within the gene to enable it to distinguish among all the genomovars of *B. cepacia*. In accordance

09763298 022004

with the present invention, it has now been determined that there is sufficient sequence variation for speciation of *B. cepacia* in the gene encoding the RecA protein (*recA*). Thus, the present invention provides a rapid molecular diagnostic test which speciates bacteria of the *B. cepacia* complex based on the detection of sequence variation within the gene encoding the RecA protein.

The sequence of the *recA* gene has been evaluated in various bacterial species and suggested as a possible basis for phylogenetic classification. Karlin et al., *J. Bacteriol.* 177: 6881-6893 (1995). One sequence of *B. cepacia* was considered by Karlin et al., but there was no discussion of speciation. A similar phylogenetic comparison was published by Eisen, J.A., *J. Mol. Evol.* 41: 1105-1123 (1995), but again no comparisons among members of the *B. cepacia* complex were made.

The sequence of *recA* genes originating from two *B. cepacia* strains have been published. Nakazawa et al., *Gene* 94: 83-88 (1990); van Waasbergen, et al., *Appl. Microbiol. Biotech.* 49: 59-65 (1998). However, beyond noting that the sequences are "highly similar" but not identical, no particular utility was suggested as a result of the differences. Surprisingly, however, even though *recA* sequences have been shown in other cases to provide results similar to 16S rRNA, sequence evaluation of the *recA* gene of sixteen strains representative of each genomovar showed that these sequences contain sufficient diversity for speciation of the *B. cepacia* complex, while the 16S rRNA does not.

The published sequences for the *recA* gene in two strains sequences of bacteria of the *B. cepacia* complex (Seq. ID Nos. 1 and 2) were used to determine two primer sequences suitable for non-specific PCR amplification of the *recA* gene. This single pair of non-degenerate primers, which have the sequences:

Forward Primer (BCR1)

TGACCGCCGAGAAGAGCAA

SEQ ID No. 3

Reverse Primer (BCR 2)

CTCTTCTTCGTCCATCGCCTC

SEQ ID No. 4

were subsequently used to amplify the *recA* gene from additional strains, and were found to be effective as amplification primers for the *recA* gene from all known members of the *B. cepacia* complex, including LMG 14191^T, the type strain for *Burkholderia pyrrocinia*.

09763298 022004

- 5 -

The amplified genes from these other strains were sequenced (Seq. ID Nos. 5-19), and the sequences aligned with the two previously known *recA* gene sequences. While there are substantial similarities, no two strains were identical except the sequences for 70431 and ATCC 17616. The *recA* genes from additional strains were sequenced and have strain numbers and GenBank Accession numbers, respectively, as follows: HI-2308, AF143777; C5424, AF143781; C1394, AF143783; ATCC 25416, AF143786; C2822, AF143792; LMG 10929, AF143793; LMG 14191 (*B. pyrrocinia*), AF143794; Ral-3, AF143795; M54, AF143796; M36, AF143797; ATCC 29424, AF143798; ATCC 53617, AF143799; ATCC 49709, AF143800; ATCC 39277, AF143801; ATCC 53266, AF143802.

The primers BCR1 and BCR2 are located at the 5'-ends of the sense and anti-sense strands, such that substantially the entire gene is amplified. Persons skilled in the art will appreciate, however, that other primers could be developed based on the sequences provided, for example to amplify only a portion of the gene in which mutations of diagnostic significance are found. Such alternative primers would be targeted to conserved regions of the *recA* gene, and could be degenerate if necessary to obtain amplification of all species within the *B. cepacia* complex.

Substantial blocks of nucleotide sequence which are conserved among all of the *recA* sequences determined are clearly from the aligned sequences. These nucleotide sequence regions may be used to design alternative primers for alternative amplification of *recA* from all bacteria of the *B. cepacia* complex. For example, there is substantial nucleotide variation from position 100 to 600 in the *recA* gene. Thus, a universal primer pair which amplified only this region would permit development of smaller amplified fragments effective for speciation. This can be accomplished using the primers:

Forward Primer (BCRU1*)

TGCGGATGGGCGACGGCG

SEQ ID No. 20

Reverse Primer (BCRU2*)

CAGTTCTGTCGCTTGATCG

SEQ ID No. 21

to produce a 485 base pair PCR product spanning a region of considerable sequence variation, or

Forward Primer (BCRU1)

ATCATGCGGATGGGCGACG

SEQ ID No. 36

09763296-022001

Reverse Primer (BCRU2)

CAGTTCTGTCGCTTGATCG

SEQ ID No. 37

to produce a 488 base pair PCR product spanning a region of considerable sequence variation.

The DNA sequence variation in the *recA* gene used to speciate the *B. cepacia* can be detected using restriction enzyme digestion, separation of the digested fragments followed by pattern matching of the resulting profile. Prior to the digestion, the *recA* gene is amplified to increase the relative abundance of this gene as compared to other nucleic acid polymers in the sample. The preferred amplification technique is PCR amplification, but other known amplification techniques may be used as well. The restriction fragment patterns contain sufficient variation due to natural variation in the 16 *recA* sequences to enable the sequence to speciate among all 5 genomovars of *B. cepacia*. The enzymes *HaeIII* and *AluI* were found to be suitably discriminatory, with *HaeIII* providing the highest degree of discrimination. Other enzymes which might possess adequate discrimination can be determined either by experimental comparison of RFLP patterns, or by computer analysis of the sequences to identify restriction sites based on the known cleavage sites of a given enzyme.

A phylogenetic tree based on the alignment of the novel sequences reported here and the two already published is shown in Figure 1 and demonstrates that the approach of the invention will clearly distinguish between all five of the current genomovars and also newly defined groups within the *B. cepacia* complex. This includes two sub-groups designated as RG-A and RG-B within genomovar III. The process of speciation of *B. cepacia* based upon RFLP using the *recA* gene scheme has been rigorously tested. It is highly discriminatory, reproducible and can be done using standard polymerase chain reaction techniques followed by agarose gel-electrophoresis.

It will be appreciated by persons skilled in the art, however, that while RFLP is a highly suitable method for evaluating nucleic acid sequence variations when (as in this case) enzymes can be identified which discriminate between the sequence variants, it is not the only method for making such a determination. Thus, nucleotide sequence information used in the method of the invention can be obtained by any other technique, including sequencing through at least the relevant regions of the *recA* gene; and the use of sequence-

09763298-022001

specific hybridization probes or ligation techniques adapted to identify sequence variations. The 485 base pair fragment generated using primers BRCU1 and BRCU2 is particularly suited for direct sequencing, since its length is amenable to complete processing in conventional commercial sequencing apparatus. The primers BRCU1 and/or BRCU2 may also be used as sequencing primers. Comparison of this 485 bp fragment of the *B. cepacia recA* gene using phylogenetic tree software and alignment software is sufficient to separate the species clusters of the *B. cepacia* complex in the same fashion as comparison of the entire gene sequence. It is, however, a more rapid approach than determination of the entire gene sequence, since only one PCR product is involved which is of an optimal size for sequencing. Genomovar-specific RFLP of this fragment can also be performed to provide speciation of *B. cepacia* complex.

A further alternative for obtaining nucleotide sequence information indicative of the genomovar to which a sample bacterium of the *B. cepacia* complex belongs is through the use of PCR primers which are genomovar-specific. Such primers are selected such that at least one primer in the pair hybridizes to a region of the *recA* gene which is not conserved, i.e., to a variable region, in such a manner that amplification only occurs if the sequence of the variable region is complementary to the primer. These specific PCR primers enable a single PCR test to be used for identification and direct detection of strains of each genomovar.

The following are non-limiting examples of genomovar-specific primers. The specificity of these primers is such that other genomovars, other than the targeted genomovar, do not produce amplification products under stringent PCR conditions.

(i) *B. multivorans* specific *recA* primers:

Forward primer (BCRBM1):

5'- CGG CGT CAA CGT GCC GGA T - 3'

SEQ ID No. 22

Reverse primer (BCRBM2):

5' - T CCA TCG CCT CGG CTT CGT - 3'

SEQ ID No. 23

PCR product expected from *B. multivorans* strains = 714 bp.

(ii) *B. vietnamiensis* specific *recA* primers:

Forward primer (BCRBV1):

- 8 -

5'- GGG CGA CGG CGA CGT GAA - 3'

SEQ ID No. 24

Reverse primer (BCRBV2):

5' - TCG GCC TTC GGC ACC AGT - 3'

SEQ ID No. 25

PCR product expected from *B. vietnamiensis* strains = 378 bp.**(iii) *B. cepacia* Genomovar IV specific *recA* primers:**

Forward primer (BCRG41):

5' - ACC GGC GAG CAG GCG CTT - 3'

SEQ ID No. 26

Reverse primer (BCRG42):

5' - ACG CCA TCG GGC ATG GCA - 3'

SEQ ID No. 27

PCR product expected from *B. cepacia* strains of genomovar IV = 647 bp.**(iv) *B. cepacia* genomovar III, RG-B *recA* specific primers.**

Forward primer:

5' - GCA AGT CAT CGC TGA GAA - 3'

SEQ ID No. 28

or

Forward primer:

5' - GCT GCA AGT CAT CGC TGA A - 3'

SEQ ID No. 38

Reverse primer):

5' - TAC GCC ATC GGG CAT GCT - 3'

SEQ ID No. 29

PCR product expected from strains of this new genomovar classification = 781 bp.

(v) *B. cepacia* genomovar I specific *recA* primers:

Forward primer (BCRG11):

5' - CAG GTC GTC TCC ACG GGT - 3'

SEQ ID No. 30

Reverse primer (BCRG12):

5' - CAC GCC GAT CTT CAT ACG A - 3'

SEQ ID No. 31

PCR product expected from strains of genomovar I = 492 bp.

(vi) *B. cepacia* genomovar III, RG-A specific *recA* primers:

Forward primer (BCRG31)

5' - GCT CGA CGT TCA ATA TGC C - 3'

SEQ ID No. 32

Reverse primer (BCRG32):

- 9 -

5' - TCG AGA CGC ACC GAC GAG - 3'

SEQ ID No. 33

PCR product expected from *B. cepacia* strains of genomovar III = 378 bp.

Additional sequencing of the complete *recA* gene from *B. cepacia* complex strains M36, M54, Ral-3 and *B. pyrrocinia* LMG 14191^T or partial sequence analysis of PCR amplicons derived from strains ATCC 29464, ATCC 53617, ATCC 39277, ATCC 49709 and ATCC 53266 was performed. The phylogenies determined using partial sequencing of this type were identical to those determined using the full sequences (Fig. 2), however, two additional clusters, denominated as RG-C and RG-D were identified. Cluster RG-C was a novel group consisting of biocontrol strains Ral-3, ATCC 53266 and M54. Comparative alignment of the *recA* sequence from Ral-3 and M54 with all other complete *B. cepacia* sequences enabled the design of RG-C specific primers with the following sequences:

Forward Primer:

GTCGGGTAAAACCGTG

SEQ ID No. 39

Reverse Primer:

TCCGCAGCCGCACCTTCA

SEQ ID No. 40

B. cepacia biocontrol strains BC-B, BC-F and AMMD all tested positive with this RG-C primer set. Thus, these primers can be used in analytical schemes for the presence of such primers, and also could be used for screening isolates for biocontrol properties.

A second novel *recA* group, RG-D, was identified which includes *B. pyrrocinia* LMG 14191^T and ATCC 32977, a strain of *B. cepacia* which produces the antibiotic xylocladin. This group is also shown in Fig. 2.

In addition to a method for identification and speciation of bacteria of the *B. cepacia* complex, the invention also provides reagents and kits suitable for carrying out this method. The reagents are generally polynucleotide primers or probes which bind to the *recA* gene of one or more strains of bacteria of the *B. cepacia* complex. One subset of the reagents of the invention are non-specific primers, such as used in Example 4 below, which are complementary to conserved regions found identically in strains of bacteria of the *B. cepacia* complex for which the sequences are given. A second subset of reagents in

- 10 -

accordance with the invention are primers/probes which can be used to selectively amplify and/or detect one genomovar of bacteria of the *B. cepacia* complex. The reagents of the invention may have a detectable or capturable label, for example a radioactive or fluorescent label or biotin, incorporated therein to facilitate evaluation of nucleotide sequence information.

Either of these types of primers/probes may be packaged in a kit with suitable reagents. These reagents may include discriminatory restriction enzymes, which are capable of producing distinctive fragment patterns to permit speciation of a bacteria-containing sample, or reagents suitable for PCR, nucleic acid sequencing and the like.

Once the species of a sample bacterium of the *B. cepacia* complex is determined using the method of the invention, it may be desirable (particularly where the bacteria is a member of an epidemic strain) to be able to provide a therapeutic agent which is effective in treating or preventing infection. Thus, the present invention further provides a vaccine composition based upon the antigenic properties of the flagellin of epidemic strains of *B. cepacia* complex for use in treating infections caused by certain species of the *B. cepacia* complex.

The use of flagellins as an antigen for vaccine purposes has been proposed in a variety of instances because of their location on the outside of bacterial cells. In the case of *B. cepacia* complex, however, Hales et al., *J. Bacteriol.* 180: 1110-1118 (1998), have reported that the flagellin gene (*fliC*) is "highly variable" and suggest its utilization as a biomarker for epidemiological and phylogenetic studies of *Burkholderia cepacia*. Such variability is inconsistent with the normal requirements that a vaccine antigen be highly conserved, such that its will be generally effective against variants of the target species. Thus, it was quite surprising to find that the subset of *B. cepacia* complex which is most transmissible have highly conserved flagellin genes which is suitable for use as a vaccine.

A total of 30 strains of bacteria of the *B. cepacia* complex were classified using the speciation method of the invention into groups based on the sequence of the *recA* gene, and were in addition characterized with respect to the BCESM (*Burkholderia cepacia* Epidemic Strain Marker) and *cbIA* markers for highly transmissible strains of *B. cepacia*. As reflected in Table 1, a substantial portion of the genomovar III strains which were positive for one or both of these markers produced a

- 11 -

single RFLP pattern (Fig. 3, pattern G) after treatment with the restriction endonuclease *HaeIII*.

Exemplary sequences and a consensus sequence for the *B. cepacia* flagellin gene, which encodes the major subunit protein of the bacterial flagellum of *B. cepacia*, have been described in the literature by Hales et al. (supra). Using the same primers described by Hales, it has been determined that the flagellin genes of *B. cepacia* strains of *recA* type III-G (genomovar III, with *recA* RFLP pattern G) are highly conserved and do not vary considerably in DNA sequence. This indicates that the protein is also highly conserved in its structure and sequence, and thus is suitable for use as an antigen for development of vaccines against the most problematic strains in patients with cystic fibrosis (CF).

In contrast, *B. multivorans* strains of *recA* type F (genomovar II), which appear less problematic in patients with CF and do not generally spread among patients, have flagellin genes which are highly variable in sequence. These data suggest that with *B. multivorans* strains, a vaccine based on the flagellum may not protect against infection with all strains types as has been the case with the bacterium *Pseudomonas aeruginosa* in CF. Thus, the observation, that the flagellin gene is actually highly conserved in the most devastating epidemic *B. cepacia* strains infecting patients with CF is apparently unique to this subset of the species of the *B. cepacia* complex.

The observation that the flagellin gene is conserved among *B. cepacia* strains which are epidemic amongst patients with CF permits the development of a vaccine based on the encoded protein antigen. The vaccine can be prepared in a variety of ways. First, protein can be purified from bacterial strains representative of this group to obtain a purified antigen. Methods for purification of flagellin from bacteria are known in the art, and can be applied to recovery of purified flagellin from epidemic strains of *B. cepacia*. Purified antigen is then used as a vaccine, with or without an adjuvant. Vaccines of this type are generally administered by subcutaneous or intramuscular injection, although other routes of administration may also be suitable. Therapeutically effective levels and frequency of vaccine administration are determined by routine monitoring of antibody titers.

In addition to the use of purified flagellin isolated directly from bacteria, it will be appreciated that the same protein, or an immunologically effective portion thereof may

- 12 -

also be prepared using, for example, recombinant technology. Thus, for example, cDNA encoding flagellin or an immunologically effective portion thereof may be cloned into a host organism and expressed to produce flagellin antigen. Smaller antigenic peptides may also be made synthetically. As used herein, the term "derived from" refers to proteins or peptides which are either isolated directly from bacteria of the *B. cepacia* complex, or which have the same amino acid sequence but which are obtained synthetically or by expression in a host organism.

Vaccine compositions in accordance with the invention comprise the flagellin or flagellin-derived antigen, in a pharmaceutically acceptable carrier. In general such carriers will be aqueous media, and may include buffers, emulsifiers, or adjuvants to enhance the immune response. Antigenic proteins or peptides may also be provided in association with lipid-carriers, e.g., liposomes or other lipid particles, in the vaccine composition.

The composition of the invention may also include other materials which are more indirectly derived from flagellin, and which provide immunoprotective therapeutic benefits. For example, human antibodies specific to *B. cepacia* flagellin or immunologically effective portion thereof are considered to be "derived-from" flagellin in accordance with the invention. Such antibodies may be administered to individuals with *B. cepacia*, genomovar III infections, to help combat the progression of infection. Such therapy is particularly suitable in end-stage *B. cepacia* infection when antibiotics and anti-inflammatory therapy have failed.

Vaccination may also be carried out using DNA vaccines of the type described generally in US Patent No. 5,580,859, which is incorporated herein by reference. DNA vaccines comprise a sequence encoding the desired protein or peptide antigen, optionally in combination with a regulatory element to control expression of the antigen. The DNA vaccine, which may be naked or incorporated in a carrier such as a liposome, is administered by subcutaneous or intramuscular injection. A "gene gun" may be used for administration.

00763298-022007

- 13 -

EXAMPLE 1

Restriction fragment length polymorphisms (RFLP) of the 16S rRNA gene amplified from strains from the *B. cepacia* complex were determined. The amplified fragment of the 16S rRNA gene was digested with the enzyme *Dde* I and the resulting products separated by agarose gel electrophoresis. Species specific patterns were found for *B. vietnamiensis* (genomovar V) and *B. multivorans* (genomovar II) strains, however sequence variation in the 16S rRNA gene was insufficient to distinguish among strains from genomovars I, III and IV.

EXAMPLE 2

The novel DNA sequences of 16 *recA* genes, from 16 strains of the genomovar representative panel shown in Table 1, have now been obtained by conventional DNA sequence analysis of portions of the *recA* gene amplified by PCR. DNA sequence analysis was performed in collaboration with Prof. Julian Davies at TerraGen Diversity Inc., Vancouver, BC, using an ABI 377 Nucleotide Sequencer. The 1040 bp *recA* gene was split into two ~520 bp PCR products to facilitate complete sequence analysis. The primer pairs used were as follows:

(i) 5' - portion of the *recA* gene:

Forward primer BCR1:

5' - TGA CCG CCG AGA AGA GCA A - 3'

SEQ ID No. 3

Reverse primer BCR4:

5' - GCG CAG CGC CTG CGA CAT - 3'

SEQ ID No. 34

These primers amplified a 527 bp product corresponding to the 5' half of the *recA* gene from all members of the *B. cepacia* complex tested. Nucleotide sequence from both strands of the amplified products was determined by direct nucleotide sequence analysis using primers BCR1 and BCR4 respectively.

(ii) 3' - portion of the *recA* gene:

Forward primer BCR 3:

5' - GTC GCA GGC GCT GCG CAA - 3'

SEQ ID No. 35

Reverse primer BCR 2:

5' - CTC TTC TTC GTC CAT CGC CTC - 3'

SEQ ID No. 4

09763298-022001
100220" 86229260

- 14 -

These primers amplified a 529 bp product corresponding to the 3' half of the *recA* gene from all members of the *B. cepacia* complex tested (see Table 1). Nucleotide sequence from both strands of the amplified products was determined by direct nucleotide sequence analysis using primers BCR2 and BCR3 respectively.

Because the amplified products described above overlapped in the regions corresponding to primers BCR3 and BCR 4, the complete 1040 bp *recA* gene sequences were derived by joining of each half of nucleotide sequence data. The alignment of these sequences was carried out in collaboration with Dr. Yossef Av-Gay, Dept. of Medicine, Vancouver General Hospital, Vancouver, BC, using the software CLUSTAL which is available on the internet. The two published *recA* sequences, b-cepd90120_1 and b-cepu70431_1 were aligned with 16 novel sequences obtained using the speciation scheme. Sequence b-cepk56-2_1 is the *recA* gene of a *B. cepacia* strain from the same cable-pilus gene encoding lineage cited in PCT Patent Applications Nos. WO 97/01647 and WO97/07237. From such a sequence alignment, the sequence variation in the *recA* gene detected by the Hae III RFLP analysis is clearly visible. These sequences can also be used for the design of genomovar-specific amplification/sequencing primers, and genomovar-specific hybridization or ligation probes.

EXAMPLE 3

A *B. cepacia* complex phylogenetic tree based on the sequence alignment of the novel *recA* genes determined was formulated as shown in Fig. 1. The alteration of single bases in the nucleotide sequence of genes by natural mutation over time can be quantitated by computer programs to an evolutionary distant which separates strains and species. Phylogenetic analysis of the *recA* sequences which I have determined clearly demonstrates that sequence variation in this gene can separate all five genomovars of the *B. cepacia* complex, shown by the tree in the figure. Even more interesting is the novel finding of two distinct groups within the strains otherwise classified within genomovar III. Thus, *recA* sequences derived from strains CEP511, C1394 and PC 184, respectively, (Table 1) cluster as a separate group on a different branch of the tree from *recA* sequences derived from strains C6433, C4455m K56-2 and C5425. These two groups have been designated a RG-B and RG-A, respectively. The three strains of the RG-B are all

- 15 -

epidemic amongst patients with CF (Mahenthiralingam et al., *J. Clin. Microbiol.* 34: 2914-2920 (1996) and encode the BCESM (Mahenthiralingam et al., *J. Clin. Microbiol.* 35: 808-816 (1997). The separate classification of these three strains based on the *recA* suggest that they may constitute a new species/genomovar group within the *B. cepacia* complex.

EXAMPLE 4

To obtain nucleotide sequence information about the *recA* genes of additional strains of bacteria of the *B. cepacia* complex (Table 1), samples of each strain were amplified using the following primers:

Forward Primer (BCR1)

TGACCGCCGAGAAGAGCAA

SEQ ID No. 3

Reverse Primer (BCR 2)

CTCTTCTTCGTCCATCGCCTC

SEQ ID No. 4

using a standard polymerase chain reaction mixture of 25 microlitres in volume (described in Mahenthiralingam et al., *J. Clin. Microbiol.* 35: 808-816 (1997)) containing 1.5 mM $MgCl_2$ and 10-20 ng of *B. cepacia* DNA. Amplification was performed as follows: 30 cycles of 1 min. at 94°C, 1 min. at 56°C, and 2 min. at 72°C, follow by a final 6 min. cycle at 72°C. This resulted in the amplification of a 1 kb DNA band corresponding to the *recA* gene of the *B. cepacia* strain being tested.

Several restriction enzyme were screened for their ability to reveal DNA sequence variation in this amplified gene which would be suitable for speciation of *B. cepacia*. The enzymes *Hae* III and *Alu* I were found to be suitably discriminatory. The restriction fragments produced by the enzyme *Hae* III were separated by agarose gel-electrophoresis, and the detected restriction fragment length polymorphisms (RFLPs) demonstrated that genomovar specific RFLPs could be generated using this approach. Representative patterns are shown in Fig. 3. (Bv = *B. vietnamiensis*, or genomovar V; Gv I = genomovar I; Bm = *B. multivorans* or genomovar II; Gv III = genomovar III and Gv IV = genomovar IV). This same approach has been applied to a panel of strains which are representative of all five genomovars of *B. cepacia* and been found to be able to distinguish strains of each genomovar (Table 1). This technique has also been applied to additional strains, and been

- 16 -

found to be reproducible and highly discriminatory for speciation of strains from the *B. cepacia* complex.

EXAMPLE 5

To obtain nucleotide sequence information about the *recA* genes of 30 strains of bacteria of the *B. cepacia* complex (Table 1), samples of each strain were amplified using the following primers:

Forward Primer (BCR1, %G/C Tm =53.2°C)

TGACCGCCGAGAAGAGCAA

SEQ ID No. 3

Reverse Primer (BCR 2, (%G/C Tm =56.3°C)

CTCTTCTTCGTCCATCGCCTC

SEQ ID No. 4

using a standard polymerase chain reaction mixture of 25 microlitres in volume (described in Mahenthiralingam et al., *J. Clin. Microbiol.* 35: 808-816 (1996)) containing 1.5 mM MgCl₂ and 10-20 ng of *B. cepacia* DNA. Amplification was performed as follows: 30 cycles of 1 min. at 94°C, 1 min. at 56°C, and 2 min. at 72°C, follow by a final 6 min. cycle at 72°C. This resulted in the amplification of a 1 kb DNA band corresponding to the *recA* gene of the *B. cepacia* strain being tested (Fig. 2A).

Several restriction enzyme were screened for their ability to reveal DNA sequence variation in this amplified gene which would be suitable for speciation of *B. cepacia*. The enzymes *Hae* III and *Alu* I were found to be suitably discriminatory. The restriction fragments produced by the enzyme *Hae* III were separated by agarose gel-electrophoresis (Fig. 2B) and the detected restriction fragment length polymorphisms (RFLPs) demonstrated that genomovar specific RFLPs could be generated using this approach (Bv = *B. vietnamiensis*, or genomovar V; Gv I = genomovar I; Bm = *B. multivorans* or genomovar II; Gv III = genomovar III and Gv IV = genomovar IV).

EXAMPLE 6

Bacterial strains identified in Example 5 as belonging to RFLP group G were epidemic strains of genomovar III which encode the BCESM DNA. When the flagellin gene of these strains was amplified using the PCR primers as described in Hales et al., a 1 kb product corresponding to the flagellin gene was obtained. Sequence variation

00763298-022001

- 17 -

in amplified flagellin gene was again detected using the restriction enzyme *Hae* III. The RFLP of the flagellin genes was highly conserved demonstrating that both the gene and the encoded protein are very conserved in sequence. The conservation in sequence of the flagellin of these epidemic strains suggests that the flagellum is an ideal candidate upon which to develop a vaccine able to protect against *B. cepacia* infection with these problematic strain types. This novel observation on the conservation of the flagellin gene and protein of epidemic *B. cepacia* strain types is in contrast to the variation reported in Hales et al. and that which is observed for *B. multivorans* strains (genomovar II) which is described in Example 7.

EXAMPLE 7

Using the same scheme outlined in Example 6, the flagellin gene in *B. multivorans* (genomovar II) strains which colonized patients attending clinic in Vancouver was examined. These strains were of a single *recA* type, BM-F, demonstrating that they are a single species type, *B. multivorans*. When the flagellin gene of these strains was examined using PCR amplification followed by RFLP analysis with *Hae* III, the gene of each of the strains examined was highly variable. This suggest that the flagellum may not be as useful in protection against infection with these strain types. However, infection with these *B. multivorans* appears not to be as problematic as infection with the epidemic *recA* III-G/BCESM positive strains.

EXAMPLE 8

Thirteen isolates having useful biological properties were evaluated to determine the *recA* group to which the isolates belong. The results of these and other tests are summarized in Table 2. Two strains within the ATCC collection with interesting catabolic properties (ATCC 29424 and ATCC 53617) were classified as *B. vietnamensis* on the basis of both 16S rRNA ARDRA and *recA* analysis. Of the remaining 11 commercially useful strains, all possessed ARDRA polymorphisms characteristic of *B. cepacia* genomovars I, III(RG-A and RG-B) and IV. Strain ATCC 49709, a seed-treatment biocontrol strain, was classified as genomovar I by *recA* RFLP and specific *recA*-PCR. Strains M36, BC-1 and BC-2, biocontrol strains isolated from the rhizosphere

09763298 "022001

- 18 -

of maize crops, possessed *recA* RFLP types characteristics of genomovar III, group RG-B and all tested positive in RG-B specific PCR. The remaining seven strains all possessed novel *recA* RFLP types and were not reactive for any of the genomovar/ sub group specific PCR primers described above.

Information of the type developed in this assay can potentially be used to assess the safety of strains of *B. cepacia* complex for commercial applications, including agricultural applications. Prior to this invention, no distinct criteria were established for division of pathogenic from non-pathogenic bacteria of the *B. cepacia* complex. However, the biologically useful strains tested were found to belong to several *recA* phylogenetic groups. Some strains cluster within the genomovar III, *recA* subgroup RG-B, which is the same subgroup to which a number of well-characterized pathogenic strains belong. This includes strain M36, which encodes the BCESM and which has been withdrawn from sale by the manufacturer. On the other hand, a number of the strains also belong to other subgroups which are not associated with pathogenicity. Thus, *recA* analysis may provide a simple method for screening biocontrol strains and other potentially useful members of the *B. cepacia* complex to identify strains that are less likely to raise health and public safety issues in their use. Furthermore, a number of the biocontrol strains were found to occupy a novel *recA* phylogenetic subgroup, RG-C. The RG-C specific PCR primers, or other tests based on the sequence analysis performed on clinical isolates indicate that RG-C strains are not encountered in human infections, and appear specifically adapted to the plant rhizosphere, they may prove to be the best template upon which to continue to development of safe biocontrol strains.

09763298-022001

TABLE 1
Provisional list of strains in the *B. cepacia* complex experimental panel

ID No.	Strain Name:	Genomovar:	Source: ^a	RAPD ^b type:	BCESM: ^c	<i>chlA</i> :	Gene RFLP type: <i>recA/Hae</i> III	16S rRNA/ <i>DdeI</i>
1	PC259	<i>B. vietnamiensis</i> (V)	CF	8	-	-	A	1
2	LMG 16232	<i>B. vietnamiensis</i> (V)	CF	X	-	-	A	1
3	ATCC 39277	<i>B. vietnamiensis</i> (V)	ENV	X	-	-	C	2
4	LMG 10929	<i>B. vietnamiensis</i> (V)	ENV	X	-	-	B	1
5	C2822	<i>B. vietnamiensis</i> (V)	CF	X	-	-	B	1
8	FC0441	<i>B. vietnamiensis</i> (V)	CGD	61	-	-	A	1
6	ATCC 25416	I	ENV	39	-	-	D	2
7	CEP0509	I	CF	41	-	+	E	2a
9	ATCC 17759	I	ENV	X	-	-	E	2a
10	C5393	<i>B. multivorans</i> (II)	CF	3	-	-	F	3
11	C3430	<i>B. multivorans</i> (II)	CF	7	-	-	F	3
12	C5274	<i>B. multivorans</i> (II)	CF	12	-	-	F	3
13	C5568	<i>B. multivorans</i> (II)	CF	19	-	-	F	3
14	249-2	<i>B. multivorans</i> (II)	LAB	30	-	-	F	3
15	ATCC 17616	<i>B. multivorans</i> (II)	ENV	30	-	-	F	3
16	LMG 13010	<i>B. multivorans</i> (II)	CF	X	-	-	F	3
17	JTC	<i>B. multivorans</i> (II)	CGD	X	-	-	F	3
18	C1257	III	CF-e	1	+	-	G	2
19	BC7	III	CF-e	2	+	+	G	2
20	J2315	III	CF-e	2	+	+	G	2
21	C5424	III	CF-e	2	+	+	G	2
22	LMG 12615	III	CF-e	2	+	+	G	2
23	C6433	III	CF-e	4	+	-	G	2
24	C4455	III	CF-e	6	+	-	G	2
25	C1394	III	CF-e	13	+	-	H	2
26	PC184	III	CF-e	17	+	-	J	2
27	CEP0511	III	CF-e	40	+	-	I	2
28	LMG 14291	IV	CF	X	-	-	J	2
29	LMG 07000	IV	CLIN	16	-	-	J	2
30	LMG 14294	IV	CF	16	-	-	J	2

Footnotes:

- ^a Source of isolate; CF, cystic fibrosis infection; CF-e, epidemic amongst CF patients (31); ENV, environment; CGD, infection of a chronic granulomatous disease patient; LAB, laboratory derived and CLIN, non-CF clinical infection.
- ^b RAPD type derived from Mahenthalingam *et al.* (30); X = unique strain RAPD fingerprint unmatched in our collection.
- ^c BCESM and *chlA* hybridization data adapted from Mahenthalingam *et al.* (31)

T00220-8629/60

Table 2
Analysis of *B. cepacia* complex strains with useful biological properties

Strain:	Source and useful property ^a :	BCESM:	16S rDNA RFLP (Dde I):	<i>recA</i> (HaeIII) RFLP:	<i>recA</i> PCR group:	Reference/ Origin
ATCC 29424	Soil isolate, capable of phthalate utilization	+	1	B	BV	51
ATCC 53617	Waste water isolate, trichloroethylene degrader	-	1	A	BV	ATCC
ATCC39277	Cornfield soil isolate, antifungal agent	-	2	P	Novel	34
ATCC 49709	Grass seed isolate, biocontrol strain	-	2	D	G1	ATCC
M36	Corn rhizosphere isolate, biocontrol strain	+	2	I	RG-B	Stine Seed Co.
BC-1 ^b	Corn rhizosphere isolate, biocontrol strain	+	2	H	RG-B	USDA
BC-2 ^b	Corn rhizosphere isolate, biocontrol strain	+	2	I	RG-B	USDA
M54	Corn rhizosphere isolate, biocontrol strain	+	2	L	RG-C	Stine Seed Co.
Ral-3	Corn rhizosphere isolate, biocontrol strain	-	2	N	RG-C	Agrium Inc.
ATCC 53266	Corn rhizosphere isolate, biocontrol strain	-	2	L	RG-C	ATCC
BC-B	Corn rhizosphere isolate, biocontrol strain	-	2	L	RG-C	USDA
BC-F ^c	Corn rhizosphere isolate, biocontrol strain	-	2	N	RG-C	USDA
AMMD	Pea rhizosphere isolate, biocontrol strain	-	2	N	RG-C	23

Footnotes:

^a Biocontrol strains demonstrated protection of crops against a various of phytopathogens; these included either fungal infection or nematode infections.

^b Strain originally isolated by K. Prakash Hebbar, USDA.

^c Strain originally isolated by WeiLi Mao, USDA.